Consensus Bond-Charge Increments Fitted to Electrostatic Potential or Field of Many Compounds: Application to MMFF94 Training Set

BRUCE L. BUSH,[1] CHRISTOPHER I. BAYLY,[2] THOMAS A. HALGREN[1]

[1] *Merck Research Laboratories, Molecular Systems Department, Rahway, New Jersey 07065*
[2] *Merck Frosst Laboratories, Department of Medicinal Chemistry, Kirkland, Quebec H9R 4P8, Canada*

Received 10 December 1998; accepted 10 May 1999

ABSTRACT: Bond-charge increments (BCIs) are additive parameters used to assign atomic charges for the MMFF force field. BCI parameters are classified parsimoniously according to two atom types and the bond order. We show how BCIs may be fitted rapidly by linear least squares to the calculated *ab initio* electrostatic potential (ESP) or to the electrostatic field. When applied simultaneously to a set of compounds or conformations, the method yields *consensus* values of the BCIs. The method can also derive conventional ''ESP-fit'' atomic charges with improved numerical stability. The method may be generalized to determine atom multipoles, multicenter charge templates, or electronegativities, but not polarizability or hardness. We determine 65 potential-derived (PD) BCI parameters, which are classified as in MMFF, by fitting the 6-31G* ESP or the electrostatic field of the 45 compounds in the original MMFF94 training set. We compare the consensus BCIs with classified BCIs that were fit to each molecule individually and with ''unique-bond'' BCIs (ESP-derived atom charges). Consensus BCIs give a satisfactory representation for about half of the structures and are robust to the adjustment of the alkyl CH bond increment to the zero value employed in MMFF94. We highlight problems at three levels: *Point approximation*: the potential near lone pairs on sulfur and to some extent nitrogen cannot be represented just by atom charges. *Bond classification*: BCIs classified according to MMFF atom types cannot represent all

This article includes Supplementary Material available from the author(s) upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/20/1495 or http://journals.wiley.com/jcc/

Correspondence to: B. L. Bush; e-mail: bruce_bush@merck.com

CCC 0192-8651 / 99 / 141495-22

delocalized electronic effects. The problem is especially severe for bonds between atoms of equivalent MMFF type, whose BCI must be taken as zero. *Consensus*: discrepancies that occur in forming the consensus across the training set indicate the need for a more detailed classification of BCIs. Contradictions are seen (e.g., between acetic acid and acetone and between guanidine and formaldehydeimine). We then test the three sets of PD-BCIs in energy minimizations of hydrogen-bonded dimers. Unique-bond BCIs used with the MMFF buffered 14−7 potential reproduce *unscaled* quantum chemical dimer interaction energies within 0.9 kcal/mol root mean square (or 0.5, omitting two N-oxides). These energies are on average 0.7 (or 0.5) kcal/mol too weak to reproduce the *scaled* quantum mechanical (SQM) results that are a benchmark for MMFF parameterization. Consensus BCIs tend to weaken the dimer energy by a further 0.4−0.6 kcal/mol. Thus, consensus PD-BCIs can serve as a starting point for MMFF parameterization, but they require both systematic and individual adjustments. Used with a ''harder'' AMBER-like Lennard−Jones potential, unique-bond PD-BCIs *without* systematic adjustment give dimer energies in fairly good agreement with SQM. J Comput Chem 20: 1495−1516, 1999

Keywords: force-field parameters; atomic charges; potential-derived charges; molecular electrostatics; consistent force field

Introduction

ATOMIC CHARGES AND CONSISTENT BOND-CHARGE INCREMENTS (BCIs)

Electrostatic interactions are crucial to a biomolecular force field. In a vacuum or nonpolar medium such interactions often dominate energy comparisons and govern conformation. In aqueous solution the force field must reflect a delicate balance of polar interactions among solutes and water molecules.

Most force fields represent the molecular charge distribution by point charges at atom centers and perhaps at lone pairs. Atom charges may be supplemented by dipoles at atoms or bonds and occasionally by quadrupoles. Occam's Razor applies: a model that is too elaborate may describe its training set precisely at the cost of generality.

Any biomolecular force field should be *conformationally transferable*. This argues for restricting the model to point charges and perhaps dipoles, because higher multipole tensors may vary strongly with the local geometry. Application to drug discovery also requires *chemically transferable* parameters that reproduce trends of binding energy, solubility, and conformation.

Atomic charges themselves are not chemically transferable parameters. Pasting together atomic charge models of separate molecules will not even conserve total molecular charge. Thus, force fields usually parameterize small templates such as functional groups and adjust at overlaps to conserve total charge.

The MMFF force field[1] was designed for routine use across the range of compounds studied in medicinal chemistry. For the sake of robustness and rapid setup, MMFF employs for its charge assignment protocol[2] the simplest possible template procedure: additive BCIs. Bonds are classified parsimoniously, chiefly according to the force-field atom types of the two bonded centers. A BCI parameter is then assigned to each class of bonds. To define atomic charges, each bond increment is added to the charge of the ''forward'' atom and subtracted from the charge of the ''back'' atom. These BCIs are similar in physical motivation to the bond dipoles used in the MM2[3] and MM3[4] programs and give almost identical results for interactions beyond about 3−4 Å in distance. BCIs are supplemented when necessary by precharges placed at ionic centers. An equivalent approach is used in the CFF force fields.[5]

For the MMFF94 release, estimates of BCI parameters were first fitted to quantum chemical calculations of the molecular dipole vector. BCIs were then adjusted in concert with other nonbond parameters to improve agreement with scaled quantum mechanical (QM) dimer geometries and energies. However, fitting just the molecular dipole

vector utilizes very little of the information available from the calculated QM electron density.

More detailed analyses of the calculated electron density fall into two broad classes: *population* methods and *electrostatic-fitting* methods. Population analyses such as Mulliken analysis[6] and ''natural'' bond orbital (NBO) analysis[7] partition the electron charge electron density from the basis set of atomic orbitals (LCAOs) directly onto atomic centers. These methods are inherently stable and give insight into chemical stability and bonding, but they are not primarily aimed at reproducing the electrostatic interactions outside the molecule and may require many multipole terms to do so.

Electrostatic potential (ESP)-fitting methods,[8–11] which are also called potential-derived (PD) charge methods, fit atomic charges (or, more generally, multicenter multipoles) to the calculated ESP (or, alternatively, the electrostatic field vector) everywhere outside the molecule. Widely used programs include PDM,[12] CHELP,[13,14] CHELPG,[15] and RESP.[16] The aim of these methods is to reproduce *interactions* between molecules with a small number of parameters.

Intermediate methods begin by determining multipoles directly from the electron density (as in population methods); then they distribute higher multipoles onto charges or lower multipoles on nearby centers. Such methods are also inherently stable. They go back at least to the work of Claverie and coworkers on multicenter multipoles[17] and include recent contributions from several laboratories.[18–21]

Instabilities in ESP-fitting methods have been noted in several works.[16,22–24] They occur especially for systems containing buried groups. Instability may lead to unphysical charges that bear no resemblance to the electron distribution and that vary widely for the same chemical fragment in different contexts. Removing independent variables can mitigate such instabilities; for example, the RESP method[16] imposes symmetry among chemically equivalent hydrogens. Restraining parameters to lie close to *a priori* values also suppresses instabilities. However, restraints and constraints degrade the representation of electrostatics.

PD charges have also been observed to vary with conformation.[23,25,26] Part of the variation is physical adjustment of the electron density, but part is an artifact of the fitting procedure. Imposing equal charges on chemically equivalent atoms (as in RESP) addresses only one aspect of this issue. Unless the force field allows conformation-dependent terms, such as polarization,[27,28] a single set of charges must be chosen to represent all conformations. A few studies[26,29–31] achieved consensus by a simultaneous fit to multiple conformations of the same compound. (Remarkably, ref. 30 includes the reaction field of a continuum solvent in determining effective solute charges.) Atoms that are buried in some conformations may be well exposed in others, allowing better determination of charges. This approach is used to generate relatively stable PD charges for the AMBER force field.[32] Our goal in the current work is to achieve a stable, transferable consensus across related chemical *compounds*, as well as *conformations*.

POTENTIAL-DERIVED BCIs

This article fits PD-BCIs simultaneously to the molecular ESPs for *all* the chemical species and conformations in a training set. BCI parameters insure charge conservation by construction. The PD-BCI method may be considered a generalization of RESP[16] in that it reduces the parameter space through chemically motivated constraints. Unlike the two-step approach of first determining PD charges and then retrofitting BCI parameters to them,[33] simultaneous fitting of BCIs is *not* a form of averaging in parameter space. Indeed, like Resat et al.,[26] we found that consensus BCIs often lie outside the range of those fit to individual structures.

The protocol described would apply to any template method that assigns atomic charges additively. ''Precharges,'' which are supplied as input to the regression, are needed for ionic centers and can also be used to represent effects that are beyond the scope of BCIs.[34]

Fitting BCI parameters to many molecules greatly reduces the number of parameters. In the application described below, 65 charge-increment classes are used to describe a training set of 45 molecular species containing 400 atoms and 367 bonds. This training set illustrates some issues of *chemical* transferability; it was not designed to explore *conformational* variation.

As noted above, PD-BCIs may be only a starting point for parameterizing atomic charges in a consistent force field. The charges must be brought into balance with the other nonbonded interactions, and the entire parameter set must be calibrated against the intended applications.

ARTICLE DESIGN

The Methods section begins with standard linear least-squares equations for fitting point charges to the ESP or the electrostatic field (''electric field''). Using a covariance-matrix (kernel) approach, we project the atom-centered matrices, quantum chemical potentials, and precharges onto BCI variables.

The Results section compares consensus BCIs, ''single-molecule'' fits, and ''unique-bond'' fits—a model equivalent to conventional ESP atom charges. We illustrate the effect of constraining the alkyl CH charge increment to zero, which was done in MMFF94 to fit torsional profiles.[35] We then test each set of PD-BCIs against the *ab initio* optimized geometry and (scaled) energy of 24 hydrogen-bonded dimers.

The Discussion addresses trade-offs: variation in parameters, quality of fit to electrostatics, and replication of quantum chemical dimers. About half of the molecules are well described by all three sets of PD-BCIs. For other molecules, buildup analysis points to the least drastic remedy:

- augmenting the atom-centered charge model (e.g., by representing selected lone pairs) as discussed in several studies,[36–38]
- generalizing BCIs to multiatom templates (e.g., to describe induction in delocalized systems), or
- adding BCI classes in a few cases of conflict between single-molecule BCIs and consensus BCIs.

BCI fits are comparatively stable, despite the use of covariance matrices, which are known to exaggerate the ''condition number.''[22]

In hydrogen-bonded dimer calculations, consensus BCIs reproduce scaled quantum mechanical (SQM) results well enough to serve as starting values for a force field, but individual adjustments are needed to attain balance; we indicate physical reasons.

We indicate in closing how the fitting protocol can determine any *additive* template of multipoles. Delocalized systems can be treated as extended templates or by incorporating quantum chemical information through precharges.[34] As shown very recently by Banks et al.,[28] similar methods can derive atom-centered polarizabilities, given assigned atom charges for each of several applied fields.

Methods: Least-Squares Fitting of Charges and BCIs

FIT TO POTENTIAL *v*

This problem has been addressed by many workers as cited above. To combine information from different molecules, we must employ the ''normal equation'' or covariance-matrix method[39] rather than singular value decomposition,[22,39] although the latter is formally more numerically stable.

We are given a set of atoms or other centers (i) and a potential v_m^{QM} at many points around the molecule. Potential v^{QM} is provided by quantum chemistry and is treated for our purposes as a set of ''observations'' to which we fit a calculated potential. The sampling points (m, ''grid points'') are typically arrayed with roughly uniform spatial density in a zone beginning just outside the region of significant electron density and extending a few Ångstroms from the molecule.

Minimize the function f as a function of the point charges q_i:

$$f = \sum_m (v_m^{\mathrm{QM}} - v_m)^2, \tag{1}$$

where the potential set up by the charges is

$$v_m = \sum_i q_i K_{im}. \tag{2}$$

Here the K_{im} are elements of the Coulomb interaction matrix, which are the ESP set up at point (m) by a unit charge at point (i) or vice versa. We choose units in which

$$K_{im} = 1/|\mathbf{d}_{im}|, \tag{3}$$

where $\mathbf{d}_{im}$ is the vector from charge point (i) to grid point (m) and $|\mathbf{d}_{im}|$ is its length. Thus, the objective function f is

$$\begin{aligned} f &= \sum_m \left(v_m^{\mathrm{QM}} - \sum_i (K_{im} q_i) \right)^2 \\ &= \sum_m \left[v_m^{\mathrm{QM}} v_m^{\mathrm{QM}} - 2 v_m^{\mathrm{QM}} \sum_i K_{im} q_i \right. \\ &\quad \left. + \sum_i \sum_j (q_i q_j K_{im} K_{jm}) \right]. \end{aligned} \tag{4}$$

At a minimum, the partial derivative of f with respect to each charge q_i is zero:

$$0 = g_i = 2\sum_j \left[q_j \sum_m K_{im} K_{jm} \right] - 2\sum_m K_{im} v_m^{\mathrm{QM}}. \quad (5)$$

Denote by C_{ij} the symmetric covariance matrix

$$C_{ij} = \sum_m K_{im} K_{jm}. \quad (6)$$

This covariance matrix contains only information about the geometry of the field points (m) and the charge points (usually atoms) (i). It contains no information on the observed potentials.

Denote by y_i the quantity

$$y_i = \sum_m K_{im} v_m^{\mathrm{QM}}. \quad (7)$$

This quantity ''maps'' the QM potential v^{QM} back onto the charge points. It encodes all the information that is needed from the quantum calculation.

Denote by f^0 the constant

$$f^0 = \sum_m v_m^{\mathrm{QM}} v_m^{\mathrm{QM}}. \quad (8)$$

Then the objective function to be minimized is

$$f = f^0 - 2\sum_i y_i q_i + \sum_i \sum_j q_i C_{ij} q_j, \quad (9)$$

and the condition at minimum is that for all charge points (i)

$$\sum_j C_{ij} q_j = y_i, \quad (10)$$

or in matrix notation

$$\boldsymbol{C}\boldsymbol{q} = \boldsymbol{y}. \quad (11)$$

This linear equation has a unique solution for the charges $\boldsymbol{q}$ if and only if matrix $\boldsymbol{C}$ can be inverted.

$$\boldsymbol{q} = \boldsymbol{C}^{-1}\boldsymbol{y} \quad (12)$$

Given a specific $\boldsymbol{y}$, one need not calculate the inverse matrix $\boldsymbol{C}^{-1}$. The solution vector $\boldsymbol{q}$ can be determined by any method for solving a linear system of equations, such as ''LU'' decomposition.[39]

The solution is necessarily a minimum (not a maximum or saddle point) of f, because this quantity is a sum of squares.

Solving $\boldsymbol{C}\boldsymbol{q} = \boldsymbol{y}$ for $\boldsymbol{q}$ will not reduce the residual f to zero. A point charge model, even with all atom charges freely determined, is incapable of reproducing the ESP set up by the actual continuous charge distribution.

Note that the summation over equispaced grid points in eq. (6) actually diverges as the first power of D_{max}, the outer ''cutoff'' distance of the sampling grid. This fact limits the physical significance of condition numbers,[14,22] whether for Coulomb matrix $\boldsymbol{K}$ or covariance $\boldsymbol{C}$.

FIELD-DERIVED CHARGES

Charges may be fitted to the electrostatic field $\mathbf{E}$ (the negative gradient of the potential v) rather than to v itself. Such charges are suitable to describe the Coulomb interaction with a probe dipole (an idealized polar functional group) rather than a monopole (a point ion).

Solving for field-derived atomic charges involves the same steps as for potential-derived charges, as outlined above. Matrix element $\mathbf{K}_{im}$ is now the electrostatic field (a spatial 3-vector) setup at point m by a unit charge at point i:

$$\mathbf{K}_{im} = \mathbf{d}_{im}/|\mathbf{d}^3|. \quad (13)$$

The covariance matrix is now a sum over grid points of dot products,

$$C_{ij} = \sum_m \mathbf{K}_{im} \cdot \mathbf{K}_{jm}. \quad (14)$$

The right-hand side of the equation involves summing the spatial dot product of $\mathbf{K}$ with the observed electrostatic field vector at each sampling point m:

$$y_i = \sum_m \mathbf{K}_{im} \cdot \mathbf{E}_m. \quad (15)$$

The residual function f now has a constant term

$$f^0 = \sum_m \mathbf{E}_m \cdot \mathbf{E}_m. \quad (16)$$

Minimizing f [eq. (9)] now achieves a match of the direction and magnitude of the field vector.

Once $\boldsymbol{C}$ and $\boldsymbol{y}$ have been calculated, the equation $\boldsymbol{C}\boldsymbol{q} = \boldsymbol{y}$ is solved for $\boldsymbol{q}$ exactly as before.

Charges fit to v and to $\mathbf{E}$ will be slightly different. (Contrary to the discussion in Chipot et al.,[37] the different solutions do not simply reflect sampling error.)

Because of the r^{-2} dependence of the field-derived **K** matrix [eq. (13)], the summations defining the field-derived covariance matrix converge, albeit slowly, as $D_{\max}^{-1}$.

FITTING BCIs AND BCI PARAMETERS

The least-squares fitting equations for the BCI parameters have the same form as those above, and the needed quantities are easily derived from $\boldsymbol{C}$ and $\boldsymbol{q}$. Suppose that all bonds of type k receive bond charge increment p_k.

Define a template matrix $\boldsymbol{T}^{\mathrm{tr}}$ with elements T_{ik} as follows. For each bond h:

1. determine forward and back atoms *if*, *ib*, and the bond type k;
2. increment the coefficient on the forward atom $T_{if,k}$ by adding $+1.0$; and
3. decrement the coefficient on the back atom $T_{ib,k}$ by adding -1.0

Atomic precharges $\boldsymbol{q}^{\mathrm{pre}}$, if any, are another necessary input to the fitting of the BCI parameters. These must be assigned during perception of atom types and bond types. The precharges for a (mono)anion sum to -1; those for a (mono)cation sum to $+1$. Precharges may be applied in other cases as well, although MMFF94 does not do so. Precharges are irrelevant to a least-squares fit when each bond or each atom is assigned a unique parameter. But within a constrained subspace of parameters, for example, the subspace spanned by a parsimonious set of BCIs, the choice of precharges will generally affect the final charge model and the residual.

Given the BCI parameters p_k and any precharge values, the atomic charges are determined as

$$q_i = q_i^{\mathrm{pre}} + \sum_k T_{ik}\, p_k. \tag{17}$$

Recall that the objective function to be minimized (as a function of atomic charges) is

$$f = f^0 - 2\sum_i y_i q_i + \sum_i \sum_j q_i C_{ij} q_j. \tag{18}$$

To express f in terms of the BCI parameters p_k, rewrite each occurrence of q_i in terms of the BCI parameters p_k and rearrange slightly. This gives

$$\begin{aligned} f = f^0 &- 2\sum_k \sum_i T_{ik}\, y_i\, p_k - 2\sum_i y_i q_i^{\mathrm{pre}} \\ &+ \sum_k \sum_l p_k \sum_i \sum_j T_{ik} C_{ij} T_{jl}\, p_l \\ &+ 2\sum_k p_k \sum_i T_{ik} C_{ij} q_j^{\mathrm{pre}}. \end{aligned} \tag{19}$$

Define a new symmetric covariance matrix $\boldsymbol{A} = \boldsymbol{T}^{\mathrm{tr}}\boldsymbol{C}\boldsymbol{T}$:

$$A_{kl} = \sum_i \sum_j T_{ik} C_{ij} T_{jl}. \tag{20}$$

Define a mapped-potential array $\mathbf{b} = \mathbf{T}^{\mathrm{tr}}\boldsymbol{y}$ and, if needed, mapped precharges $\boldsymbol{T}^{\mathrm{tr}}\boldsymbol{C}\boldsymbol{q}^{\mathrm{pre}}$:

$$b_k = \sum_i T_{ik}\, y_i + \sum_i \sum_j T_{ik} C_{ij} q_j^{\mathrm{pre}}. \tag{21}$$

Finally, if there are precharges, define a constant $z^{\mathrm{pre}} = \boldsymbol{y}^{\mathrm{tr}}\boldsymbol{q}^{\mathrm{pre}}$:

$$z^{\mathrm{pre}} = \sum_i y_i q_i^{\mathrm{pre}}. \tag{22}$$

Note that $\boldsymbol{A}$, $\boldsymbol{b}$ for the BCI parameters can be calculated directly from the quantities $\boldsymbol{C}$, $\boldsymbol{y}$ used in fitting atomic charges. Even when exploring different rules for defining precharges, there is no need to recalculate potentials or to adjust $\boldsymbol{C}$.

The objective function [eq. (19)] becomes

$$f = (f^0 - 2z^{\mathrm{pre}}) - 2\sum_k b_k\, p_k + \sum_k \sum_l p_k\, A_{kl}\, p_l. \tag{23}$$

The system of equations for fitting the parameters p_k by minimizing f is formally the same as before:

$$\sum_l A_{kl}\, p_l = b_k \tag{24}$$

As previously, this can be solved by inverting **A** or by solving $\boldsymbol{A}\boldsymbol{p} = \boldsymbol{b}$ through Gaussian elimination.

Fitting a parsimonious set of BCIs generally leads to different atomic charges and a larger residual f than does an ''all-atom'' fit. However, the BCI formalism can reproduce the results of an all-atom fit constrained only by total molecular charge. Simply assign to each bond its own BCI parameter, omitting one bond in each ring.

Interestingly, using BCIs rather than monopoles as variables insures that the covariance matrix converges at long distances; therefore, compar-

isons of the condition number for the BCI formalism (in the Discussion below) become well defined. Elements of $\boldsymbol{A}$ converge as $D_{\max}^{-1}$ for fits to the potential and as $D_{\max}^{-3}$ for fits to the field.

SIMULTANEOUS FITS TO MANY STRUCTURES

A major advantage of using a common set of BCI parameters is that these BCIs may be fit simultaneously to many structures s (chemical compounds or conformations of each compound). Sum the matrices and vectors over all structures, then solve the usual equations:

$$\boldsymbol{A}^{\text{tot}}\boldsymbol{p} = \boldsymbol{b}^{\text{tot}},$$

where

$$\begin{aligned} \boldsymbol{A}^{\text{tot}} &= \sum_s \boldsymbol{A}^s, \\ \boldsymbol{b}^{\text{tot}} &= \sum_s \boldsymbol{b}^s. \end{aligned} \tag{25}$$

The equations minimize the squared residual summed over all points around all the structures.

EIGENANALYSIS AND STABILITY

Stability is always a concern when solving a matrix equation of the form $\boldsymbol{Ap} = \boldsymbol{b}$. If each element of $\boldsymbol{b}$ is subject to random independent "noise", this noise is multiplied by the ratio $(a_{\min}/a_{\max})$ of the smallest to the largest eigenvalue of $\boldsymbol{A}$. This quantity, the inverse *condition number*, multiplies any roundoff error imposed by machine numerical precision, but it has limited chemical significance. Another measure of instability is $1/a_{\min}$, which multiplies any chance correlations between the parameters that enter into the least well-determined eigenvector.

Constraining a solution by reducing the number of parameters or restraining parameters by adding to f a term proportional to $(\boldsymbol{p} - \boldsymbol{p}^*)^2$ will always improve "stability." The price, of course, is that either method will degrade the precision of fit (increase the residual of f).

We calculate $a_{\max}$ and $a_{\min}$, the largest and smallest eigenvalues of $\boldsymbol{A}$, by an adaptation of the classic *power method*.[40] Although inefficient, this is easily fast enough for our applications and can be adapted transparently to find eigenvalues (and vectors) for $a_{\min}$, as well as $a_{\max}$. Trial vector $\boldsymbol{t}$ is initialized to the actual mapped-potential vector $\boldsymbol{b}$. Upon repeated normalization to unit length and multiplication (or division) by $\boldsymbol{A}$, it converges to the eigenvector associated with the largest (smallest) eigenvalue of $\boldsymbol{A}$.

To find $a_{\max}$:

$$\begin{aligned} &\boldsymbol{t}_0 = \boldsymbol{b}; \quad \boldsymbol{u}_n = \boldsymbol{t}_n/|\boldsymbol{t}_n|; \quad \boldsymbol{t}_{n+1} = \boldsymbol{A}\boldsymbol{u}_n; \\ &a_{\max} = \text{limit } (n \text{ large})[(\boldsymbol{t}_{n+1} \cdot \boldsymbol{u}_n)]. \end{aligned} \tag{26}$$

To find $a_{\min}$:

$$\begin{aligned} &\boldsymbol{t}_0 = \boldsymbol{b}; \quad \boldsymbol{u}_n = \boldsymbol{t}_n/|\boldsymbol{t}_n|; \quad \boldsymbol{t}_{n+1} = \boldsymbol{A}^{-1}\boldsymbol{u}_n; \\ &a_{\min} = \text{limit } (n \text{ large})[1/(\boldsymbol{t}_{n+1} \cdot \boldsymbol{u}_n)]. \end{aligned} \tag{27}$$

This procedure remains within a subspace relevant to the solution of $\boldsymbol{Ax} = \boldsymbol{b}$. We accelerate convergence with a Jacobi rotation[40] so that each $\boldsymbol{t}_{n+1}$ is the exact eigenvector of $\boldsymbol{A}$ within the 2-dimensional subspace spanned by $\boldsymbol{At}_n$ and $\boldsymbol{A}(\boldsymbol{At}_n)$, and convergence is governed by the ratio of the *third* largest to the largest eigenvalue of $\boldsymbol{A}$ (or of $\boldsymbol{A}^{-1}$).

Fitting all structures simultaneously by summing $\boldsymbol{A}_s$ insures that the least stable eigenvalue $(a_{\min})$ is governed by the best-determined examples, which is in accord with the intuitive notion that well-planned simultaneous fits improve stability.

Software Design

Unix routines combined as "filters" perform a postanalysis of output from existing software. All information on parameter types comes from the OPTIMOL[1] program in MMFF mode. Independent of molecular geometry, it defines template matrix $\mathbf{T}$ and the atom precharges $\boldsymbol{q}^{\text{pre}}$ for each structure. From the Gaussian 92[41] quantum chemistry program, filtered through a potential-mapping procedure, come the atom-based covariance matrix $\boldsymbol{C}$ and the mapped-potential vector $\boldsymbol{y}$ for each structure. This information is the same as that used by RESP. All information on molecular geometry, basis set, and spatial sampling of the field comes from this file.

The program *make_bmx* combines the atom-based matrix $\boldsymbol{C}$ and the sampled potential array $\boldsymbol{v}$ with template matrix $\boldsymbol{T}$ and precharges $\boldsymbol{q}^{\text{pre}}$ to generate $\boldsymbol{A}$ and $\boldsymbol{b}$. Program *sum_bmx* sums covariance matrices $\boldsymbol{A}$ and mapped-potential vectors

$\boldsymbol{b}$ over a set of compounds and/or conformations. Finally, program *solve_bmx* solves ($\boldsymbol{Ap} = \boldsymbol{b}$) for parameters $\boldsymbol{p}$ and reports residuals, eigenvalues, and eigenvectors. It can use a file of parameters $\boldsymbol{p}^*$ as constraints or restraints.

Application to 45 Molecules in MMFF94 Training Set

The training set of 45 structures, which were chosen to exemplify 65 BCI parameter types judged most important for describing intermolecular interactions, was used in the refinement[2] of nonbond parameters during the original development of MMFF94. (A much larger set of compounds and conformations was used in the initial estimation of MMFF parameter values.) Conformational variations are not explored in this training set, and relatively few chemical contexts of each bond type are represented.

Compounds and BCI parameter types are shown in Figure 1. Conformation codes are as in ref. 1. Parameter codes represent relative row numbers of the bond-stretching parameters in the distributed MMFF94 parameter file.[1,2] One bond in each set of equivalent bonds is labeled toward its forward atom. Precharges on ionic centers are shown in italics in the figure. Note especially the highlighted symmetric bonds between atoms of the same MMFF atom type. In the current BCI protocol these must receive a bond charge increment of zero. This is a very stringent, and in some cases costly, constraint on representing electrostatics. More detailed information, including a key to conformation codes and the initial value of residual f^0, is tabulated in Table A-1 of the Supplementary Material.

Quantum calculations were performed on each molecule with the Gaussian 92 program.[41] Geometries were minimized at the Hartree–Fock (HF)/6-31G* level. We used the same van der Waals radii as in ref. 9 (H, 1.20; C, 1.50; N, 1.50; O, 1.40; F, 1.35; P, 1.80; S, 1.75; Cl, 1.70) using a ''face-centered cubic'' sampling method in the surface zone lying between 1.4 and 2.0 times the atom radius (C. I. Bayly, unpublished procedure). Each surface atom has approximately one sampling point per 0.09 Å^2 at the midplane of the sampled zone; this density varies by only 1–2%. Potentials and electrostatic field vectors were calculated at these sampling points, using the keywords ''prop = (grid,field) nosymm'' in Gaussian 92.[41]

An inadvertent error in sampling confirmed the robustness of fitting BCIs simultaneously to the training set. Molecule AR02a (pyridine) was originally sampled at only 1 point/0.28 Å^2 and was later corrected to standard density. Consensus charges changed by < 0.007 units for the BCI types directly involved and typically changed by 0.001 units for other BCI parameters; consensus residuals changed imperceptibly. Charges fit to the potential of AR02a alone changed by < 0.01; and to field by < 0.012; relative root mean square (RMS) residuals changed only from 19.6 to 20.4% (potential fit) and from 35.0 to 37.5% (field fit).

The various BCI models were compared for their ability to reproduce SQM interaction energies and geometries. Hydrogen-bonded dimers constructed from this training set, many involving a water molecule as a partner, were geometry optimized by quantum chemistry (HF/6-31G*) and by molecular mechanics. Several force fields were constructed. Consensus, single-molecule, and unique-bond BCI charges (equivalent to ESP-fit atomic charges) and the BCI parameters employed in MMFF94 were combined with the MMFF buffered 14–7 van der Waals potential. Additionally, the unique-bond BCI charges were combined with an AMBER-like Lennard–Jones (LJ) 12–6 potential with AMBER 95[32] radii and well depths mapped to MMFF atom types insofar as possible. [A hybrid van der Waals radius was used for divalent oxygen (1.70 Å) as a compromise between AMBER alcohol and ether oxygen radii (1.721 and 1.684 Å, respectively). Thus, interaction energies for dimers involving hydrogen bonds to divalent oxygen are expected to be larger in magnitude for hydroxyls and smaller for ethers than those using true AMBER 95.[32]] Drawings of the dimer geometries were given previously.[2] As described elsewhere,[42] *scaled* QM interaction energies are HF/6-31G* values increased by 15% for the water dimer and by 10% for other neutral dimers. *Scaled* X⋯Z heteroatom distances (used in comparisons summarized in a later table) are HF/6-31G* values reduced by 8%.

Results and Discussion

We first compare sets of BCI parameters, and then discuss the residuals. Each step of the consensus is examined: all-atom (unique-bond) fit, the best (single-molecule) fit allowed by the BCI classification, and lastly the consensus over the training

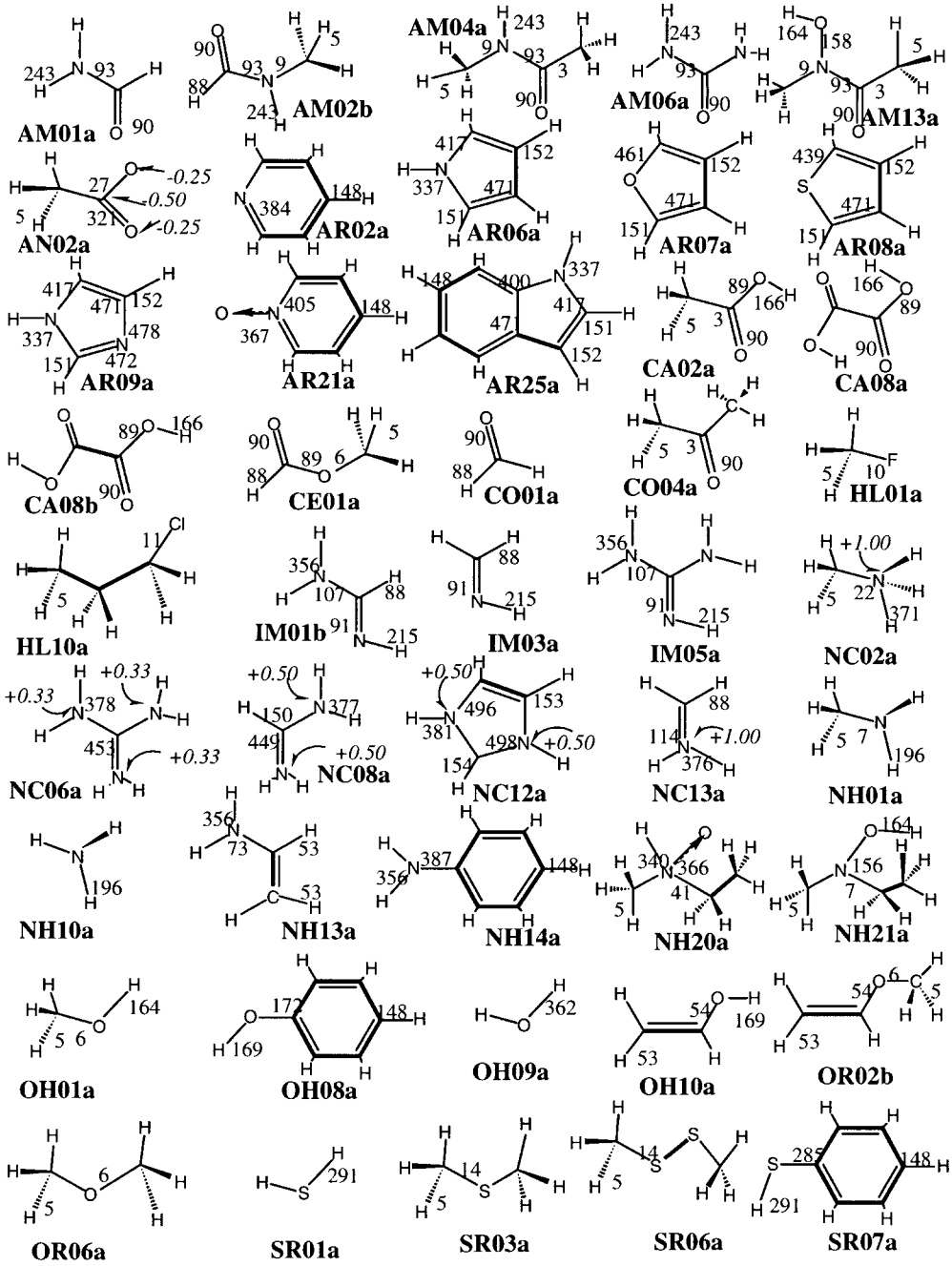

FIGURE 1. Training set compounds with bond-charge increment codes.

set. We reexamine molecules that have large residuals or extreme values of BCI parameters. We conclude that the transferability of BCI parameters is governed by chemical factors. We identify chemical functions poorly represented by the current BCI classification and discuss appropriate improvements. Consensus BCIs are then evaluated with respect to their ability to represent hydrogen-bonded dimers.

COMPARISON OF BCI PARAMETERS: FITS TO POTENTIAL AND FIELD

The consensus BCI parameters fit to ESP or to an electrostatic field are shown in Table I. MMFF94 parameter values are included for reference. As may be seen from Figure 2, the BCI parameters are quite similar whether they are fit to the potential v or the electrostatic field $\mathbf{E}$. This agreement applies to individual molecules, not just to the consensus

TABLE I.
Consensus BCI Parameters Fitted to Electrostatic Potential and to Electrostatic Field Compared with MMFF94 Values.

Parm.[a] No.	A Types[b]	ESP Fit		FLD Fit		
		ESP Fit	p(CH) = 0	FLD Fit	p(CH) = 0	MMFF94
3	CC0103	0.0133	−0.0498	0.0224	−0.0437	−0.0610
5	CH0105	0.0758	0.0000	0.0793	0.0000	0
6	CO0106	−0.1982	−0.2577	−0.2116	−0.2721	−0.2800
7	CN0108	−0.1631	−0.2251	−0.1849	−0.2496	−0.2700
9	CN0110	−0.2026	−0.2679	−0.2192	−0.2868	−0.3001
10	CF0111	−0.2459	−0.3032	−0.2546	−0.3098	−0.3400
11	CC0112	−0.2552	−0.2884	−0.2732	−0.3006	−0.2900
14	CS0115	−0.1406	−0.1846	−0.1516	−0.1981	−0.2300
22	CN0134	−0.3507	−0.4074	−0.3755	−0.4352	−0.5030
27	CC0141	0.2690	0.2066	0.2593	0.1978	0.1060
41	CN0168	−0.1007	−0.1383	−0.1181	−0.1639	−0.2560
53	CH0205	0.1470	0.1476	0.1457	0.1455	0.1500
54	CO0206	−0.0344	−0.0350	−0.0616	−0.0652	−0.0767
73	CN0240	−0.0132	−0.0125	−0.0532	−0.0531	−0.1000
88	CH0305	−0.0217	−0.0153	0.0061	0.0108	0.0600
89	CO0306	−0.1518	−0.1493	−0.1713	−0.1700	−0.1500
90	CO0307	−0.5408	−0.5367	−0.5596	−0.5555	−0.5700
91	CN0309	−0.4503	−0.4457	−0.4545	−0.4518	−0.4500
93	CN0310	−0.0836	−0.0843	−0.1067	−0.1090	−0.0600
107	CN0340	−0.1074	−0.1040	−0.1364	−0.1346	−0.0500
114	CN0354	−0.9504	−0.9420	−0.8660	−0.8610	−0.4000
148	HC0537	−0.1295	−0.1295	−0.1317	−0.1317	−0.1500
150	HC0557	−0.1270	−0.1270	−0.1278	−0.1278	−0.1500
151	HC0563	−0.1873	−0.1873	−0.1661	−0.1661	−0.1500
152	HC0564	−0.1513	−0.1513	−0.1442	−0.1442	−0.1500
153	HC0578	−0.2528	−0.2528	−0.2368	−0.2368	−0.1500
154	HC0580	−0.2377	−0.2377	−0.2205	−0.2205	−0.1500
156	ON0608	0.0405	0.0401	0.0514	0.0511	−0.1000
158	ON0610	0.0275	0.0254	0.0458	0.0454	0.0355
164	OH0621	0.3900	0.3939	0.4187	0.4238	0.4000
166	OH0624	0.4314	0.4340	0.4516	0.4540	0.5000
169	OH0629	0.4112	0.4112	0.4144	0.4148	0.4500
172	OC0637	0.0695	0.0695	0.1243	0.1243	0.0825
196	NH0823	0.3579	0.3578	0.3704	0.3719	0.3600
215	NH0927	0.3611	0.3604	0.3594	0.3588	0.4000
243	NH1028	0.3671	0.3698	0.3771	0.3803	0.3700
285	SC1537	0.1314	0.1314	0.1384	0.1384	0.1015
291	SH1571	0.1730	0.1731	0.1938	0.1938	0.1800
337	HN2339	−0.2953	−0.2953	−0.3119	−0.3119	−0.2700
340	HN2368	−0.1834	−0.1497	−0.2373	−0.2078	−0.3600
356	HN2840	−0.3669	−0.3663	−0.3813	−0.3810	−0.4000
362	HO3170	−0.4078	−0.4078	−0.4208	−0.4208	−0.4300
363	OC3241	0.6166	0.6016	0.6265	0.6107	0.6500
366	ON3268	0.6257	0.6415	0.6633	0.6749	0.7500
367	ON3269	0.5418	0.5418	0.6083	0.6083	0.7500
371	NH3436	0.3271	0.3382	0.3713	0.3848	0.4500
376	HN3654	−0.5880	−0.5837	−0.5279	−0.5257	−0.4000
377	HN3655	−0.4808	−0.4808	−0.4750	−0.4750	−0.4500
378	HN3656	−0.5075	−0.5075	−0.5084	−0.5084	−0.4500

(Continued)

TABLE I.
(Continued)

Parm.[a] No.	A Types[b]	ESP Fit		FLD Fit		
		ESP Fit	p(CH) = 0	FLD Fit	p(CH) = 0	MMFF94
381	HN3681	−0.3695	−0.3695	−0.3696	−0.3696	−0.4500
384	CN3738	−0.2389	−0.2389	−0.2287	−0.2287	−0.3100
387	CN3740	−0.0390	−0.0387	−0.1060	−0.1058	−0.1000
400	CC3763	0.0622	0.0622	0.0828	0.0828	0
402	CC3764	0.0533	0.0533	0.0559	0.0559	0
405	CN3769	−0.1900	−0.1899	−0.1287	−0.1287	−0.0895
417	NC3963	−0.0795	−0.0795	−0.0347	−0.0347	−0.1516
439	SC4463	0.0177	0.0177	0.0246	0.0246	0.0400
449	NC5557	0.4218	0.4218	0.4305	0.4305	0.3544
453	NC5657	0.4196	0.4196	0.4423	0.4423	0.4000
461	OC5963	0.0729	0.0729	0.0899	0.0899	0.1400
471	CC6364	−0.0151	−0.0151	−0.0294	−0.0294	0
472	CN6366	−0.2571	−0.2571	−0.2915	−0.2915	−0.3381
478	CN6466	−0.1717	−0.1717	−0.1727	−0.1727	−0.2272
496	CN7881	−0.1215	−0.1215	−0.1371	−0.1371	−0.3500
498	CN8081	−0.1534	−0.1534	−0.1744	−0.1744	−0.4000

[a] The parameter numbers (Parm. No.) are relative line numbers of bond-stretching parameters in the distributed MMFF94 parameter file.[1, 2]

[b] *XYnnmm* designates a BCI parameter between an atom *X* of numerical MMFF atom type *nn* and an atom *Y* of MMFF atom type *mm*, where *X* and *Y* are ''back'' and ''forward'' atoms, respectively.

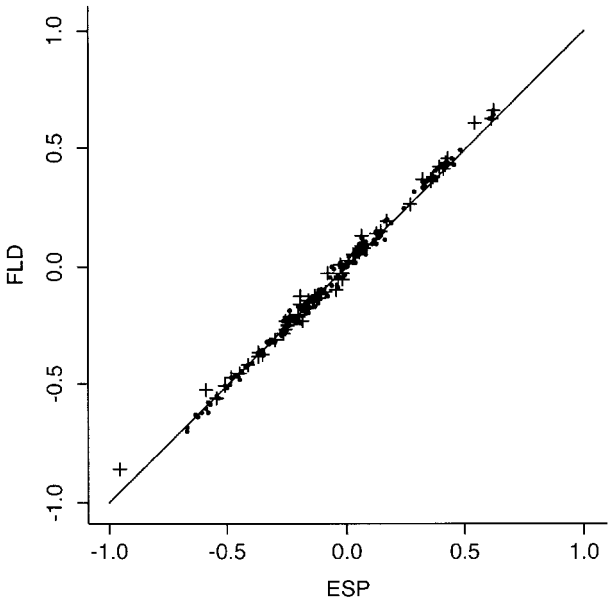

FIGURE 2. Comparison of BCI parameters fitted to potential (ESP) and to field (FLD). Consensus values are displayed as pluses and single-molecule values are points.

(see Table A-II in Supplementary Material). BCIs fit to the ESP and to the field vector appear to be equally satisfactory. (The condition number for the fit to the field vector is more stable, but we show that all BCI fits are so stable that the marginal improvement is unimportant.) Because potential-fit values are prevalent in the literature, we emphasize these values below; conclusions for field-fit values are very similar.

EFFECT OF CONSTRAINING THE ALKYL CH BOND INCREMENT

Setting BCI parameter 5 (alkyl CH) to zero rather than its consensus potential-fit value (+0.076) is essentially the same as adopting a united-atom approximation for alkyl electrostatics. Table I indicates that this constraint causes concerted, compensating changes of 0.076 or less in the consensus values of the 10 other parameters that involve the alkyl carbon (MMFF atom type 1). Other parameters change only imperceptibly. Considerable cancellation of dipoles occurs as CH bonds are summed vectorially, and further cancellation occurs as the consensus is formed between compounds.

This stability is not a trivial result. In their study[9] of PD atomic charges, Singh and Kollman found that for some structures like 1-methyluracil, changing from an all-atom to a united-atom representation caused changes in other atom charges of up to 0.6 electron units. Moreover, the resultant united-atom charges did not correspond even approximately to folding the hydrogen charges into those of the parent atoms. The use of classified BCIs apparently suppresses correlations of charge separation along parallel bonds, resulting in a far more stable and interpretable model.

VARIATION OF SINGLE-MOLECULE BCI PARAMETERS

Although quite robust to the effect of changing one parameter, BCI parameters can vary by over 0.20 charge units when fitted to different molecules. Figure 3 illustrates this variability for potential fits. Numerical values for potential and field fits are tabulated in the Supplementary Material (Table A-II).

This variability would be benign if the consensus BCIs reproduced the potentials almost as well as those fit to individual molecules. In fact, the consensus proved quite satisfactory in about half of the molecules. In other cases we systematically identified the level of consensus at which parameters become unable to reproduce the potentials.

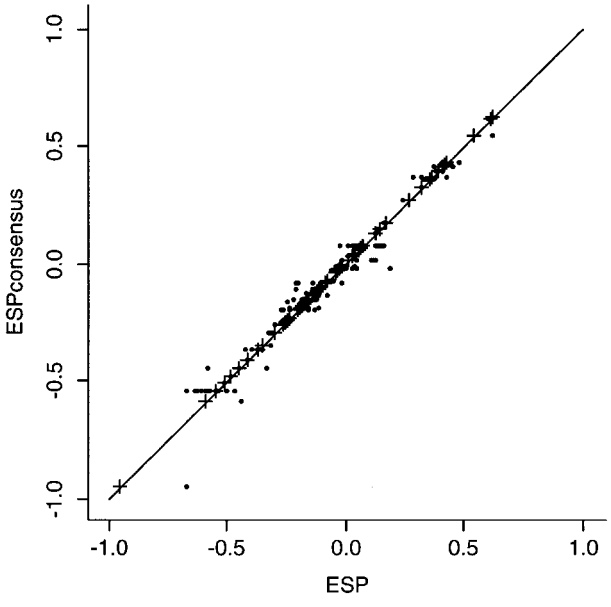

FIGURE 3. BCI parameters fitted to electrostatic potential: consensus and individual-molecule values. Consensus values are plotted as plus signs and individual-molecule values are points. The vertical distribution shows the variability of any one parameter.

RESIDUALS FOR CONSENSUS FIT AND FOR SINGLE-MOLECULE FITS

Table II gives the molecule by molecule residuals for fits to the ESP. We show a series of regressions at increasing levels of classification and consensus:

- unique bond: every bond has its own increment (omitting one bond per ring). This gives exactly the same charges and residuals as a conventional ESP-fit atomic charge constrained by total molecular charge. Such atom charges are independent of any distribution of precharges.
- single mol: the BCI classification used by the MMFF force field is applied and fitted separately to each molecule. In this case and the remaining two, any change to the precharging template generally affects the charges and residuals.
- consens: classified BCIs are fitted simultaneously to all 45 molecules to yield a consensus.
- consens, p(CH) = 0: consensus with the BCI for alkyl CH held to a value of 0.

Many of the molecules are satisfactorily fitted at all levels of consensus, even with zero BCI on alkyl CH. We take a residual f of 0.0100 as a rough benchmark for ''satisfactory'' (in the arbitrary units used here). For methanol (OH01a) such an error would approximately double the 11.6% relative RMS error of a unique-bond (all-atom) ESP fit, which is slightly more than the effect of using consensus BCIs.

The residual f for the consensus ESP fit is less than this benchmark for more than half of the training set (23 molecules), even though most molecules have a much larger surface area than methanol. For these molecules it is reasonable to expect that consensus BCIs, balanced with other nonbonded terms, may allow calculation of hydrogen-bond dimer energies within a range of 0.4–0.9 kcal/mol. This range can be taken as a goal for the *chemical* transferability of the BCIs.

At the other extreme, sulfur-containing compounds (aliphatics SR01a, SR03a, and SR07a and, to some extent, the aromatic thiophene AR08a) are poorly described even by a unique-bond (all-atom)

TABLE II.
Residuals for Bond-Charge Increment Fits to the Electrostatic Potential.

	Absolute Residual				Relative RMS			
Mol Code	Unique Bond	Single Mol	Consens	Consens p(CH) = 0	Unique Bond	Single Mol	Consens	Consens p(CH) = 0
AM01a	0.0037	0.0043	0.0089[a]	0.0086	0.0686	0.0733	0.1060	0.104
AM02b	0.0047	0.0070	0.0104	0.0098	0.0805	0.0981	0.1200	0.116
AM04a	0.0035	0.0051	0.0094[a]	0.0123	0.0687	0.0838	0.1130	0.130
AM06a	0.0021	0.0022	0.0089[a]	0.0091	0.0507	0.0523	0.1050	0.106
AM13a	0.0089	0.0117	0.0176[a]	0.0205	0.1250	0.1440	0.1760	0.190
AN02a	0.0016	0.0017	0.0019	0.0020	0.0756	0.0774	0.0819	0.084
AR02a	0.0054	0.0149[b]	0.0170	0.0170	0.1260	0.2040	0.2250	0.225
AR06a	0.0036[a]	0.0036	0.0050[a]	0.0050	0.0988	0.0988	0.1160	0.116
AR07a	0.0078	0.0078	0.0083	0.0083	0.2190	0.2190	0.2250	0.225
AR08a	0.0113[a]	0.0113	0.0136	0.0136	0.2470	0.2470	0.2720	0.272
AR09a	0.0101[a]	0.0115	0.0140	0.0140	0.1160	0.1240	0.1370	0.137
AR21a	0.0026	0.0036	0.0090	0.0090	0.0493	0.0584	0.0922	0.092
AR25a	0.0027	0.0083	0.0087	0.0087	0.0740	0.1300	0.1330	0.133
CA02a	0.0014	0.0017	0.0129[b]	0.0133	0.0633	0.0694	0.1940	0.196
CA08a	0.0030	0.0030	0.0054	0.0051	0.0808	0.0807	0.1080	0.105
CA08b	0.0022	0.0022	0.0065[a]	0.0060	0.0593	0.0592	0.1030	0.099
CE01a	0.0061	0.0061	0.0145[b]	0.0149	0.1360	0.1360	0.2090	0.212
CO01a	0.0014	0.0014	0.0070[a]	0.0072	0.0590	0.0590	0.1320	0.134
CO04a	0.0008	0.0009	0.0067[a]	0.0102[b]	0.0411	0.0445	0.1190	0.148
HL01a	0.0010	0.0010	0.0011	0.0016	0.0678	0.0678	0.0710	0.085
HL10a	0.0083[a]	0.0149	0.0153	0.0159	0.1760	0.2360	0.2400	0.244
IM01b	0.0043	0.0049	0.0098[a]	0.0095	0.0881	0.0940	0.1330	0.131
Im03a	0.0036	0.0048	0.0147[b]	0.0144	0.1010	0.1170	0.2050	0.203
IM05a	0.0030	0.0043	0.0128[b]	0.0132	0.0678	0.0811	0.1400	0.142
NC02a	0.0035	0.0035	0.0038	0.0058	0.2060	0.2050	0.2150	0.264
NC06a	0.0013	0.0012	0.0012	0.0012	0.0524	0.0501	0.0501	0.050
NC08a	0.0015	0.0029	0.0029	0.0029	0.0693	0.0954	0.0954	0.095
NC12a	0.0036	0.0036	0.0036	0.0036	0.0862	0.0861	0.0861	0.086
NC13a	0.0015	0.0014	0.0143[b]	0.0135	0.0515	0.0511	0.1620	0.157
NH01a	0.0090[a]	0.0127	0.0152	0.0133	0.2030	0.2400	0.2630	0.246
NH10a	0.0055	0.0055	0.0059	0.0059	0.1310	0.1310	0.1360	0.136
NH13a	0.0064	0.0399[b]	0.0402	0.0401	0.1540	0.3830	0.3850	0.385
NH14a	0.0050	0.0183[b]	0.0198	0.0198	0.1120	0.2150	0.2230	0.223
NH20a	0.0034	0.0146[b]	0.0146	0.0177	0.0526	0.1090	0.1090	0.120
NH21a	0.0065	0.0091	0.0097	0.0124	0.2020	0.2390	0.2450	0.278
OH01a	0.0039	0.0075	0.0083	0.0078	0.1160	0.1610	0.1700	0.164
OH08a	0.0024	0.0139[b]	0.0155	0.0155	0.0826	0.1980	0.2090	0.209
OH09a	0.0049	0.0050	0.0050	0.0050	0.1120	0.1120	0.1120	0.112
OH10a	0.0032	0.0130[b]	0.0152	0.0152	0.0936	0.1890	0.2050	0.205
OR02b	0.0032	0.0118[b]	0.0133	0.0129	0.1140	0.2180	0.2310	0.228
OR06a	0.0038	0.0047	0.0105[b]	0.0118	0.1450	0.1610	0.2410	0.256
SR01a	0.0332[b]	0.0332	0.0333	0.0333	0.4340	0.4340	0.4350	0.435
SR03a	0.0204[b]	0.0220	0.0259	0.0330	0.3000	0.3110	0.3380	0.381
SR06a	0.0209[b]	0.0234	0.0259	0.0316	0.2650	0.2800	0.2940	0.326
SR07a	0.0181[b]	0.0207	0.0209	0.0209	0.2700	0.2890	0.2900	0.290

The relative RMS is the RMS residual, $\sqrt{(f/f^0)}$. The values with different letter superscripts indicate an [a] large or [b] very large difference from the preceding column (from 0.0 for the unique-bond column). See Table A-III in Supplementary Material for residuals in fitting the field rather than the potential.

fit. Residuals f for the aliphatics are about 0.020–0.030 in our units. Nonetheless (see below), energies of dimers involving these compounds are in error by only 0.3–0.6 kcal/mol compared to SQM dimer energies. This is true when SH is a donor and when S is an acceptor, so the good results are likely not due to a fortunate choice of nonbond radii. Sampling the fit potential as close as 2.5 Å from sulfur may exaggerate nearby errors compared to those involved in typical nonbonded contacts at distances of > 3.0 Å. Neglected anisotropies may also tend to cancel at hydrogen-bond geometries.

Imidazole (AR09a) and furan (AR07a) display a similar problem to a lesser extent. Pyrrole (AR06a) is free of this problem, indicating that the in-plane lone pair (not out-of-plane density) is responsible. The effects on dimer energies (see below) are substantial and are in the range of 0.6–1.4 kcal/mol. This problem is neither improved nor made worse by classifying the BCIs or forming a consensus.

To describe these molecules with reasonable accuracy requires going beyond atom-centered charges. Dinur and Hagler[43] showed that the addition of atom-centered dipoles and, for second-row atoms like sulfur, quadrupoles, markedly improves the fit to the ESP. Similarly, Dixon and Kollman[38] demonstrated significant improvements upon adding lone pairs to selected heteroatoms. Defining BCIs on lone-pair ''bonds'' would have the same effect.

MOLECULES POORLY REPRESENTED BY CLASSIFIED BCIs

For about 15 of the molecules the atom-monopole representation is adequate, but the MMFF BCI classification degrades the quality of fit con-

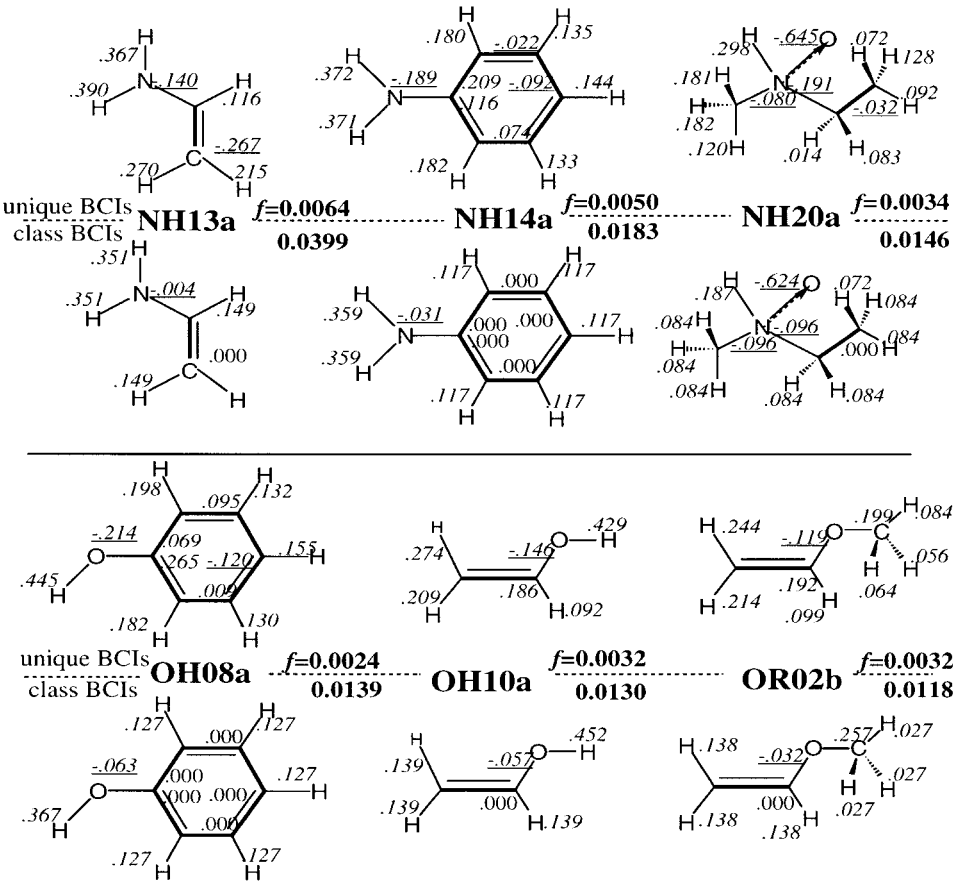

FIGURE 4. Structures for which the MMFF BCI parameter classification substantially degrades the fit to the electrostatic potential. ''Class BCIs'' in this comparison has the same meaning as ''single mol'' in Figure 5.

siderably (see Fig. 4 and the highlighted entries in the ''single mol'' column of Table II). The worst cases involve π-resonance effects in molecules that contain locally symmetric bonds, such as vinylic or aromatic C=C bonds.

These resonant systems represent a fundamental challenge to any classification scheme. Templates of more than two bonded atoms will be needed. A template might move charge between 1,3 atoms in an aromatic ring or even between an exocyclic substituent and the ortho and para ring positions. It would be very difficult to control the number of parameters for large templates. We are therefore pursuing an alternative: combining atom population charges from QM population analysis with a small number of parameterized BCIs.[34]

CARBOXY/KETO COMPOUNDS AND UNSATURATED AMINES

Problems that occur at the level of forming a consensus across the training set are less fundamental (see highlighted entries in the ''consens'' column of Table II). These can be addressed by reclassifying BCIs or adding new BCI classes. A systematic buildup analysis is needed to judge the

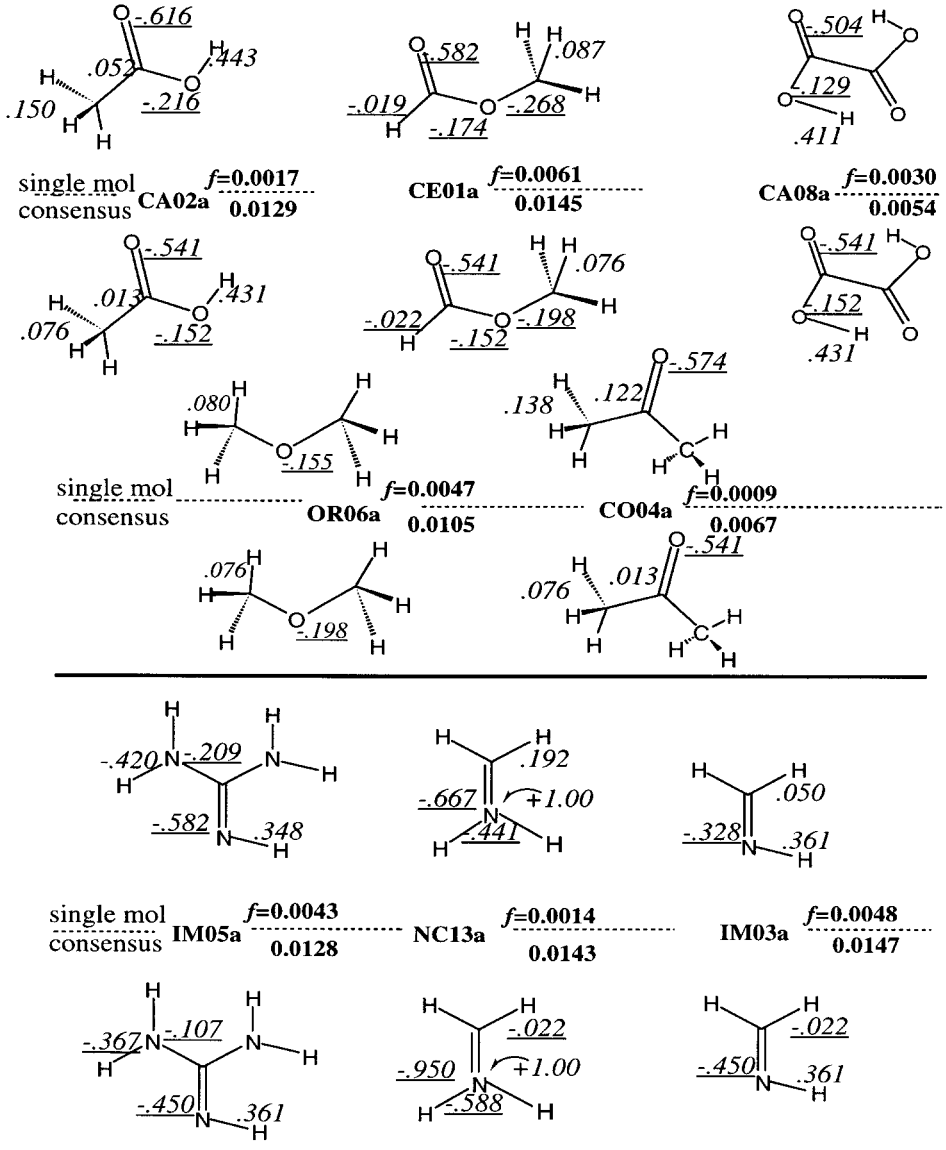

FIGURE 5. Structures for which using consensus BCI parameters degrades the fit to the electrostatic potential.

best compromise between parsimony and precision, which is required by Occam's Razor. Figure 5 illustrates two interesting sets of contradictions involving oxygen and nitrogen atom types.

Carboxy and Keto Compounds

Acetic acid (CA02a) and methyl formate (CE01a) can be well fit by a consensus between the two but not by parameters fitted simultaneously to the full training set. The extreme contrast of BCI parameters between these compounds and internally hydrogen-bonded oxalic acid (CA08a) seemed to offer an explanation at first.

In fact, however, CA08a has relatively little ''leverage'' on the consensus. We tested every combination of all eight training compounds that contain the two oxygen atom types, MMFF types 7 (carbonyl) and 6 (divalent ether, alcohol, ester, or acid). (The 256 combinations require only 2–3 min on an SGI workstation.) Table III shows the effect of adding one compound at a time in the order that least degrades the residual f (error of fit) in representing three ''subject'' molecules: acid

TABLE III.
Residuals for Buildup to Consensus: Acetic Acid, Methyl Formate, and Dimethyl Ether.

ca02a	ca08a	ca08b	ce01a	co01a	co04a	oh01a	or06a	*N*
				Acetic Acid (CA02a)				
0.0017	0	0	0	0	0	0	0	1
0.0020	0	0	0	0	0	0	0.0048	2
0.0025	0	0	0	0	0	0.0089	0.0052	3
0.0035	0	0	0.0160	0	0	0.0079	0.0084	4
0.0047	0	0.0031	0.0159	0	0	0.0078	0.0084	5
0.0056	0.0056	0.0048	0.0156	0	0	0.0078	0.0082	6
0.0062	0.0049	0.0051	0.0159	0.0034	0	0.0079	0.0079	7
0.0131	0.0041	0.0068	0.0160	0.0033	**0.0075**	0.0079	0.0079	8
				Methyl Formate (CE01a)				
0	0	0	0.0061	0	0	0	0	1
0	0	0	0.0063	0	0.0010	0	0	2
0.0041	0	0.0035	0.0062	0	0	0	0	3
0.0122	0	0.0052	0.0066	0	0.0075	0	0	4
0.0125	0.0052	0.0065	0.0069	0	0.0071	0	0	5
0.0129	0.0045	0.0067	0.0091	0.0043	0.0071	0	0	6
0.0130	0.0042	0.0067	**0.0116**	0.0036	0.0074	**0.0094**	0	7
0.0131	0.0041	0.0068	**0.0160**	0.0033	0.0075	0.0079	**0.0079**	8
				Dimethyl Ether (OR06a)				
0	0	0	0	0	0	0	0.0047	1
0	0	0	0	0.0014	0	0	0.0047	2
0	0	0	0	0.0026	0.0026	0	0.0047	3
0	0.0030	0	0	0.0024	0.0028	0	0.0047	4
0	0.0049	0.0050	0	0.0029	0.0022	0	0.0047	5
0.0131	0.0043	0.0068	0	0.0032	0.0072	0	0.0047	6
0.0131	0.0042	0.0067	0	0.0032	0.0075	0.0087	0.0053	7
0.0131	0.0041	0.0068	**0.0160**	0.0033	0.0075	0.0079	**0.0079**	8

The BCIs are fit to the electrostatic potential of the subject molecule alone, then other structures are added in the order of the least increase to the residual of this subject molecule. Each row shows residuals for fitting BCIs simultaneously to the *N* structures whose entries are nonzero. Bold entries highlight the structures whose addition greatly degrades the subject residual. See the text.

CA02a, ester CE01a, and ether OR06a. Clearly, *disparity in fit parameters is not a reliable guide to problems in forming a consensus*. The fit to:

- acid CA02a is most degraded by acetone CO04a and vice versa (data not shown for the latter).
- ester CE01a is most degraded by ether OR06a and vice versa. The ester is also affected by methanol OH01a, although these share only H_3C—O parameters (5 and 6).
- formaldehyde CO01a is little affected by consensus with the other structures or vice versa, despite its electronically anomalous carbonyl (a charge increment of only 0.46).

By contrast, the BCIs for the two variants of internally hydrogen-bonded *trans*-oxalic acid (CA08a and CA08b) are not affected by simultaneous fit with the other structures or vice versa. This result may reflect their $\mathbf{C}_{2h}$ symmetry and zero molecular dipole moment.

The individually fitted CA08a and CA08b charge parameters are puzzling in themselves. One would expect internal hydrogen bonding to polarize the C—O and the O—H bonds, but the charge increments are *smaller* not *larger* than usual. Apparently, potential sampling ''sees'' a blurred combination of electron clouds on terminal =O and —H atoms, and the fit is a compromise that gives smaller than expected charge increments. This is a warning that *PD charges may not give a reasonable picture of internal charge redistribution* any more than atom monopoles from population analysis can represent the external ESP.

To achieve a consensus BCI model for these oxygen-containing compounds may require only an added lone-pair increment. Failing this, it may require separate BCI parameters for ketone and acid carbonyls and for ester, ether, acid, and alcohol C—O bonds.

Nonaliphatic Amines

Buildup analysis identifies problems at several levels. The fit to cation NC13a is badly degraded by IM03a. This problem is relatively easily solved by either placing substantial positive precharges on the CH_2 hydogens (data not shown) or, perhaps better, assigning a new BCI classification to their bonds. A more fundamental problem is that neutral amines IM03a and IM05a cannot be reconciled, because of the inability of this BCI classification to represent resonance among the CN bonds of guanidine (IM05a). Lone-pair bond increments are not likely to improve the consensus in this case; consensus and all-atom fits shift charge only slightly (0.02 |e|) along the vector parallel to the missing lone pair.

CONDITION NUMBERS AND STABILITY

Reducing the number of variables must improve (decrease) the inverse condition number (a_{max}/a_{min}). For a fit to the ESP the value is reduced by over 2 orders of magnitude by simply using unique-bond increments as variables. For AR25a it decreases from 304,000 to 1873; more typical values are 10,000 and 500. Numerical values for all levels of consensus are tabulated in the Supplementary Material, Table A-IV.

This improvement in condition number is caused not by constraining the total charge as such, but by the method employed. In fact, using Lagrange multipliers to constrain charge usually *increases* the inverse condition number slightly (data not shown), confirming a conjecture of Sigfridsson and Ryde.[22]

Classifying BCIs by MMFF parameter type further reduces the number of variables and improves the condition number. In most cases, fitting to the electrostatic field rather than to the potential brings slight further improvement (see FLD_BCI vs. ESP_BCI of Table A-IV in the Supplementary Material).

As a measure of instability we prefer ($1/a_{min}$), the dominant eigenvalue of the inverse matrix. This depends on the degree of ''burial'' and the local geometry of functional groups. Unlike a_{max}, it is insensitive to purely numerical choices such as the outer sampling radius and the method used to constrain the molecular charge and/or dipole. The corresponding eigenvector indicates parameters that are the most correlated and thus the most sensitive and probably the least transferable. Transforming atom charge parameters into unique-bond increments improves this measure of stability only modestly (2–5-fold). But classifying BCIs brings larger improvements in some cases, even when fitting a single molecule (ESP_BCI vs. ESP−bnd in Table A-IV). This shows the value of using *classified* BCIs in cases where some instances appear in an exposed or well-determined context. Consensus between molecules brings further improvement in many cases, but not in all. In our

relatively nonredundant training set, the dominant $(1/a_{\min})$ is nearly as large (24.1) for the consensus fit as it is for the ''worst'' molecule, N-oxide NH20a. This happens because only one molecule determines the least stable combination of parameters $(0.47^{*}p_{CN0168} + 0.70^{*}p_{HN2368} + 0.54^{*}p_{ON3268})$ around the tetrahedral nitrogen. Notably, this molecule is a bad outlier in the assessment of hydrogen-bonded dimers (see below).

REPRODUCTION OF INTERMOLECULAR INTERACTION ENERGIES

Consensus and single-molecule potential-fit BCI parameters and unique-bond BCIs (ESP-fit atom charges) were used to optimize dimer interaction energies, and they were tested by comparing the force-field interaction energies and geometries with both unscaled QM and SQM results. The scaling is designed to account for omission of many-body terms such as polarization of solutes in aqueous solution. We employed the 24 dimers used in a recent assessment of several widely available force fields.[42] Detailed comparisons for the full training set of 66 small-molecule dimers[2] are provided in the Supplementary Material, Tables B-I through B-IV.

Table IV gives the results for individual dimers. Rows are grouped to highlight questions of *balance*, which is the correct ranking of comparable dimers. Table V summarizes the statistics. To separate the issues of *scale* and *balance*, we tabulated the systematic error (average signed error), RMS deviation from this systematic offset, and RMS

TABLE IV.
Dimer Energies at Optimized Geometry for BCI Methods Compared with Unscaled Quantum Mechanical (QM) and Scaled Quantum Mechanical (SQM 6-31G*) Energies.

Dimer	Monomer	QM	SQM	MMFF 94	Consensus BCI, B-14–7	Single-Mol BCI, B-14–7	Unique-Bond BCI	
							B-14–7	LJ-12–6
$HOH\cdots OH_2$	OH09a	−5.62	−6.47	−6.61	−5.72	−5.72	**−5.71**	**−6.74**
$CH_3OH\cdots OHCH_3$	OH01a	−5.36	−6.09	−6.03	−5.31	−5.37	**−6.62**	**−7.79**
$CH_3OH\cdots OH_2$	OH01a	−5.59	−6.15	−6.14	−5.49	−5.58	−6.37	−7.55
$HOH\cdots OHCH_3$	OH01a	−5.55	−6.10	−6.32	−5.42	−5.36	−5.85	−6.81
$C_6H_5OH\cdots OH_2$	OH08a	−7.35	−8.09	−8.45	−7.06	−6.33	−7.38	−8.55
$HOH\cdots OHC_6H_5$	OH08a	−4.69	−5.16	−5.08	−4.22	−3.87	−5.01	−6.21
HOH···*t*-NMA	AM04a	−7.30	−8.03	−8.19	−7.01	−7.37	−7.50	−8.49
t-NMA$\cdots OH_2$	AM04a	−5.43	−5.97	−6.86	−6.28	−5.61	−5.90	−6.43
$HOH\cdots O{=}CHOCH_3$	CE01a	−5.81	−6.39	−6.84	−5.92	−6.65	−6.67	−7.79
$HOH\cdots O(CH_3)CH{=}O$	CE01a	−3.47	−3.82	−2.63	−1.90	−2.56	−2.56	−3.29
Acetic acid, cyclic	CA02a	−15.55	−17.10	−17.10	−13.12	−15.19	−15.14	−16.92
Formamide, cyclic	AM01a	−13.44	−14.78	−12.26	−10.68	−13.04	−13.98	−15.07
$HOH\cdots NH_3$	NH10a	−6.56	−7.22	−6.83	−6.28	−6.63	−6.61	−7.71
HOH···pyridine	AR02a	−6.03	−6.63	−7.10	−4.99	−5.03	−5.63	−6.20
HOH···thiophene	AR08a	−2.42	−2.66	−2.71	−2.83	−2.95	−2.95	−4.50
HOH···imidazole	AR09a	−7.06	−7.77	−7.94	**−5.86**	**−6.15**	−6.31	**−6.64**
Imidazole$\cdots OH_2$	AR09a	−6.36	−7.00	−7.92	**−6.48**	**−6.42**	−6.46	**−7.11**
$C_6H_5NH_2\cdots OH_2$	NH14a	−4.20	−4.62	−4.86	−4.45	−4.40	−4.46	−5.10
$NH_3\cdots O(CH_3)_2$	NH10a, OR06a	−2.92	−3.21	−3.34	−3.03	−2.65	−2.76	−3.40
$HOH\cdots MeEtNH{-}O$	NH20a	−14.07	−15.48	−14.87	−9.21	−9.22	−10.53	−12.03
$HOH\cdots C_5H_5N{-}O$	AR21a	−9.72	−10.69	−10.61	−7.75	−8.34	−8.23	−9.95
$HOH\cdots S(CH_3)_2$	SR03a	−3.24	−3.56	−3.48	−2.81	−3.06	−3.04	−4.10
$HSH\cdots OH_2$	SR01a	−2.66	−2.93	−2.91	−2.61	−2.67	−2.68	−3.47
$HSH\cdots SH_2$	SR01a	−0.87	−0.96	−1.25	−1.17	−1.21	−1.21	−1.59

Underlined values differ from the reference SQM value by 2 kcal / mol or more; values in **bold** disagree with the SQM values by 1 kcal / mol or more. The QM and SQM are the unscaled and scaled (HF / 6-31G*) energies of dimerization (kcal / mol), respectively, at optimal QM geometry. See the text. B-14–7 is the buffered 14–7 van der Waals potential used with MMFF94; LJ-12–6 is a Lennard–Jones 12–6 potential similar to that used in AMBER. See the text.

TABLE V.
Statistics of Force-Field Optimized Energies and Geometries for BCI Methods Compared with Unscaled (QM) and Scaled Quantum Mechanical (SQM) Results.

	Interaction Energy (kcal / mol)				X ··· Z (Å)		X — H ··· Z
BCI Method, Nonbond[a] Potential	RMS Dev.	X RMS to Avg.	Avg. Dev.	Max. Dev.	RMS Dev.	Avg. Dev.	RMS Dev.
Vs. Unscaled QM, 24 dimers							
MMFF94	0.88	0.62	−0.62	1.54	0.22	−0.21	16.8°
Consensus BCI, Buf-14−7	1.41	1.25	0.66	4.86	0.19	−0.15	19.3°
Single-molecule BCI, Buf-14−7	1.15	1.07	0.42	4.85	0.20	−0.15	21.3°
Unique-bond BCI, Buf-14−7	0.92	0.92	0.08	3.54	0.21	−0.18	21.5°
Unique-bond BCI, LJ-12−6	1.29	0.92	−0.90	2.26	0.28	−0.24	20.2°
Vs. SQM, 24 dimers							
MMFF94	0.67	0.67	0.02	2.52	0.09	0.04	16.8°
Consensus BCI, Buf-14−7	2.02	1.54	1.30	6.27	0.15	0.11	19.3°
Single-molecule BCI, Buf-14−7	1.67	1.29	1.06	6.26	0.15	0.10	21.3°
Unique-bond BCI, Buf-14−7	1.34	1.13	0.72	4.96	0.13	0.08	21.5°
Unique-bond BCI, LJ-12−6	1.09	1.06	−0.26	3.45	0.13	0.01	20.2°
Vs. Unscaled QM, 22 dimers							
MMFF94	0.89	0.66	−0.60	1.54	0.23	−0.21	17.3°
Consensus BCI, Buf-14−7	0.96	0.87	0.41	2.75	0.19	−0.16	19.7°
Single-molecule BCI, Buf-14−7	0.53	0.49	0.17	1.04	0.20	−0.17	21.9°
Unique-bond BCI, Buf-14−7	0.50	0.48	−0.14	1.09	0.22	−0.19	22.3°
Unique-bond BCI, LJ-12−6	1.28	0.70	−1.07	2.26	0.30	−0.26	20.6°
Vs. SQM, 22 dimers							
MMFF94	0.69	0.69	−0.01	2.52	0.09	0.04	17.3°
Consensus BCI, Buf-14−7	1.51	1.13	1.00	4.10	0.13	0.09	19.7°
Single-molecule BCI, Buf-14−7	1.01	0.65	0.77	2.35	0.13	0.08	21.9°
Unique-bond BCI, Buf-14−7	0.75	0.60	0.45	1.96	0.12	0.06	22.3°
Unique-bond BCI, LJ-12−6	0.84	0.69	−0.49	1.84	0.11	−0.01	20.6°

[a]The nonbond potential is the MMFF buffered 14−7 (Buf-14−7) unless indicated as LJ-12−6. See the text. RMS Dev., root mean square deviation from the average; Avg. Dev., average signed deviation.

deviation from the QM value, which combines both effects. Table V also gives statistics that omit the N-oxides NH20a and AR21a. (These compounds are outliers for the PD-BCI models, although they are well fit by MMFF94 BCI parameters.)

Addressing first the issue of *scale* among the force-field models:

- Consensus BCIs give slightly weaker dimers (less negative interaction energies, longer nonbonded distances) than BCIs fit to individual molecules, and these in turn give weaker dimers than unique-bond BCIs. This trend spans a range of about 0.5–0.6 kcal/mol.
- The AMBER-like LJ 12–6 potential lowers energies (strengthens dimers) by about 1.0 kcal/mol compared with the MMFF buffered 14–7 potential.

Comparing the force-field models and the SQM results over the reduced set of 22 dimers:

- Unique-bond BCIs used with the MMFF 14–7 potential give interaction energies that are on average 0.45 kcal/mol too weak compared with SQM. Energies for classified BCIs are weaker yet: 0.77 and 1.00 kcal/mol for the single-molecule and consensus BCIs compared to SQM.
- Unique-bond BCIs used with the AMBER-like LJ 12–6 potential give energies that are slightly (0.3–0.5 kcal/mol) too strong on average compared to SQM. Average hydrogen-bonded (X ··· Z) distances agree well with SQM. It seems likely, therefore, that in practical applications classified BCIs would give results that *on average* are suited for use with an AMBER-like LJ 12–6 potential.

MMFF94 still fits the SQM energies, X ⋯ Z heteroatom distances, and X—H ⋯ Z hydrogen bond angles of this training better than does any combination of van der Waals models with PD-BCIs, including unique-bond BCIs.

- PD-BCIs coupled with the MMFF buffered 14−7 potential do better in reproducing *unscaled* rather than *scaled* QM results. In all cases, the most precisely fit (unique bond) give the best results.

The systematic effect of the nonbond potential is easily understood. The softer MMFF buffered 14−7 potential stores more repulsive energy at the energy-optimized dimer geometry, hence it requires charges that are enhanced (by about 5%, data not shown) to give a more negative electrostatic interaction energy. The harder AMBER-based LJ 12−6 potential requires no change in scale to fit the same target data.

PD-BCIs also require adjustment to improve *balance*. For example, they sometimes predict methanol to be an appreciably stronger hydrogen-bond donor and acceptor than water. (However, the great stability of the methanol dimer in the LJ 12−6 calculation may be mainly due to the hybrid radius used for divalent oxygen.) In addition, all combinations of van der Waals models with PD-BCIs predict that imidazole is a stronger hydrogen-bond acceptor than donor; and all underestimate, sometimes severely, the acceptor strengths of pyridine and the two amine oxides.

Adjustments to charges and to the nonbond model serve to compensate for physical effects omitted from the force-field calculation but included implicitly in the SQM results. The effects of lone-pair anisotropy on hydrogen-bond directionality are well recognized, but the effects on distance and direction may be less well appreciated. Standard molecular mechanics provides no minimum in the ESP in the lone-pair region. As a result, a stronger van der Waals repulsion is required to ward off ''nuclear collapse,'' while a more negative atom-centered acceptor charge is needed to reproduce the greater stability and closer approach of the SQM dimer. Other neglected phenomena include intermolecular polarization and charge transfer.

In summary, systematically rescaled consensus PD-BCIs appear to be a reasonable starting point for parameterizing a consistent force field. Further modest adjustments, which correct a scatter of about 1 kcal/mol, will be required to give an appropriate rank order to competing hydrogen bonds. If BCIs are used with an AMBER-like nonbond potential, systematic rescaling of dipoles is probably not necessary, but such rescaling is required for use with the softer buffered 14−7 potential used in MMFF.

GENERALIZATIONS OF BCI METHOD

The method illustrated here can easily be generalized to any *additive* template of point charges or point multipoles. To fit multipoint templates one uses the same covariance matrix $\boldsymbol{C}$ but a different template matrix $\boldsymbol{T}$ [eq. (20)]. To fit multipoles such as atom- or bond-centered dipoles, one must redefine $\boldsymbol{C}$ from appropriate Green's functions analogous to the Coulomb matrices in eqs. (3) and (13). For example, to fit a dipole at point (i) along unit vector $\mathbf{u}$ requires a Coulomb matrix element

$$K_{im} = (\mathbf{u} \cdot \mathbf{d}_{im})/d_{im}^3. \qquad (28)$$

Linear regression cannot be used to determine parameters that define the potential (or field) *nonadditively*. It can determine electronegativities, which are linearly related to charge shifts, but not polarizabilities or hardnesses, which enter quadratically. However, approximate fitting of some of these parameters is possible in the sense of linearized small-signal theory or perturbation theory, as was done in work reported very recently.[28]

Conclusions

The covariance-matrix formalism allows simultaneous fitting of BCI parameters to external ESPs of multiple structures. The numerical stability of BCI fits is governed by chemical factors, not by round-off or sampling error. Disparity in BCI parameters (or atom charges) does not identify the sources of problems in achieving consensus; systematic buildup analysis is needed.

Consensus BCIs classified according to the MMFF94 prescription give a good fit to ESP for half of the 45 molecules in the MMFF training set. About six molecules may need lone pairs or atom-centered dipoles or quadrupoles to represent the potential around sulfur or nitrogen. Additional BCI classes may be needed for acids and car-

bonyls. Deeper failures of the BCI classification occur around locally symmetric bonds, especially in π-delocalized systems. Our future reports will show that precharging schemes based on quantum chemical electron populations can address some of these problems.[34]

Consensus PD-BCIs appear to be good starting points for parameterizing a consistent, transferable force field. Tests on hydrogen-bonded dimers indicate that, as reported in ref. 32, PD-BCIs may be suitable for use ''as is'' in a force field such as AMBER, giving results comparable to those of SQM. By contrast, the softer buffered 14–7 nonbond potential used in MMFF necessitates a general increase in the scale of the BCIs. With either nonbond model, further adjustment of individual PD-BCIs to correct dimer energies by about ± 1 kcal/mol is needed to *balance* competing polar interactions.

Supplementary Materials

Our Supplementary Materials contain details of the monomer training set, values of BCIs fit to individual molecules, residuals for fits to the electrostatic field, stability analysis of potential and field fits, and comparisons between force-field and QM results for the full 66-dimer training set.

References

1. Halgren, T. A. J Comput Chem 1996, 17, 490–519.
2. Halgren, T. A. J Comput Chem 1996, 17, 520–552.
3. Allinger, N. L. J Am Chem Soc 1997, 99, 8127.
4. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. J Am Chem Soc 1989, 111, 8551, 8566, 8576.
5. Maple, J. R.; Hwang, M.-J.; Stockfish, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. J Comput Chem 1994, 15, 161–182.
6. Mulliken, R. S. J Chem Phys 1955, 23, 1833, 1841.
7. Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. In Frisch, M. J.; Trucks, G. W.; Head–Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian 92, Revision C; Gaussian, Inc.: Pittsburgh, PA, 1992; Link 607, NBO version 3.1.
8. Kollman, P. A. J Am Chem Soc 1977, 99, 4875.
9. Singh, U. C.; Kollman, P. A. J Comput Chem 1984, 5, 129–145.
10. Luque, F. J.; Illas, F.; Orozco, F. J Comput Chem 1990, 11, 416.
11. Ferenczy, G. G.; Reynolds, C. A.; Richards, W. G. J Comput Chem 1990, 11, 159.
12. Cox, S. R.; Williams, D. E. J Comput Chem 1981, 2, 304.
13. Chirlian, L. E.; Francl, M. M. QCPE Bull 1987, 7, 39.
14. Chirlian, L. E.; Francl, M. M. J Comput Chem 1987, 8, 894–905.
15. Breneman, C.; Wiberg, K. B. J Comput Chem 1990, 11, 361–373.
16. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269.
17. (a) Claverie, P. In Intermolecular Interactions: From Diatomics to Biopolymers; Pullman, B., ed.; Wiley: New York, 1978; Chapter 2; (b) Vigné-Maeder, F.; Claverie, P. J Chem Phys 1988, 88, 4934–4948.
18. Thole, B. T.; VanDuijnen, P. T. Theor Chim Acta 1983, 63, 209.
19. Ritchie, J. P.; Copenhaver, A. S. J Comput Chem 1995, 16, 777–789.
20. Winn, P. J.; Ferenczy, G. G.; Reynolds, C. A. J Phys Chem A 1997, 101, 5437–5445.
21. Sokalski, W. A.; Shibata, M.; Ornstein, R. L.; Rein, R. Theor Chim Acta 1993, 85, 209–221.
22. Sigfridsson, E.; Ryde, U. J Comput Chem 1998, 19, 377–395.
23. Stouch, T. R.; Williams, D. E. J Comput Chem 1992, 13, 622–632.
24. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. J Comput Chem 1996, 17, 367–383.
25. Williams, D. E. Biopolymers 1990, 29, 1367.
26. Resat, H.; Maye, P. V.; Mezei, M. Biopolymers 1997, 41, 73–81.
27. Dinur, U.; Hagler, A. T. J Comput Chem 1995, 16, 154.
28. Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. J Chem Phys 1999, 110, 741–754.
29. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. J Comput Chem 1995, 16, 1357.
30. Gabdoulline, R. R.; Wade, R. C. J Phys Chem 1996, 100, 3868–3878.
31. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. J Am Chem Soc 1993, 115, 9620–9631.
32. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179–5197.
33. Sun, H. S.; Mumbry, S. J.; Maple, J. R.; Hagler, A. T. J Am Chem Soc 1994, 116, 2978–2987.
34. Jakalian, A.; Bayly, C. I.; Halgren, T. A.; Bush, B. L. unpublished data.
35. Halgren, T. A.; Nachbar, R. B. J Comput Chem 1996, 17, 587–615.
36. Colonna, F.; Angyan, J. G.; Tapia, O. Chem Phys Lett 1990, 172, 55.
37. Chipot, C.; Maigret, B.; Rivail, J. L.; Scheraga, H. A. J Phys Chem 1992, 96, 10276–10284.
38. Dixon, R. W.; Kollman, P. A. J Comput Chem 1997, 18, 1632–1646.

39. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vettering, W. T. Numerical Recipes; Cambridge University Press: New York, 1986.
40. Mathews, J. H. Numerical Methods for Mathematics, Science, Engineering, 2nd ed.; Prentice−Hall: Englewood Cliffs, NJ, 1992; Sections 11.2, 11.3.
41. Frisch, M. J.; Trucks, G. W.; Head−Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian 92, Revision C; Gaussian, Inc.: Pittsburgh, PA, 1992.
42. Halgren, T. A. J Comput Chem 1999, 20, 730−748.
43. (a) Dinur, U.; Hagler, A. T. J Chem Phys 1989, 91, 2949−2958; (b) Dinur, U.; Hagler, A. T. J Chem Phys 1989, 91, 2959−2970.